DOCUMENT RESUME

ED 303 723 CG 021 421

AUTHOR Baehr, Melany E.
TITLE Criteria of Police Officer Performance.
PUB DATE Aug 88
NOTE 17p.; Paper presented at the Annual Meeting of the American Psychological Association (96th, Atlanta, GA, August 12-16, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Evaluation Criteria; *Evaluation Methods; *Performance; *Performance Factors; *Personnel Evaluation; *Police

ABSTRACT

The indifferent success in achieving the goal of reliable and valid police officer work performance measures has been attributed to the complexity of the job. This study explored reasons for unsatisfactory police performance evaluation, reviewing and integrating previous studies and new recently collected but not yet published data. Performance criteria studied include subjective supervisory assessments and several objective criteria such as arrests, commendations and disciplinary actions. A quantitative job analysis at three different tenure points was implemented. The job analysis survey (Job Functions Inventory for Police Officers) was completed on a volunteer basis by police officers (N=37) in two mid-sized police departments. The three length of service groupings were 0-6 years, 8-16 years, and 18 years or more. A discriminant function analysis produced four bipolar functions. The two poles of each function were treated as separate factors and thus produced eight scoring dimensions for the Inventory. Each of the eight dimensions had a counterpart among the eight dimensions identified in a 1968 study (Baehr, Furcon, Froemel). A possible imaginative description of these results showed the low-tenure officers engaged in street patrol, and active and hazardous criminal investigation activities. The latter activities continued in the middle-tenure group which also undertook special promotion and training assignments. Finally, the high-tenure group was chiefly concerned with enforcing police procedures and regulations and carrying out public duties. (ABL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED303723

CRITERIA OF POLICE OFFICER PERFORMANCE*

Melany E. Baehr, Senior Consultant

London House, Inc., Park Ridge, IL

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Melany E. Baehr

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

* Presented at the Symposium "Psychological Contributions to Police Selection and Assessment," 1988 APA Annual Convention, Atlanta, GA.

SUMMARY

Criteria of Police Performance *

Melany E. Baehr, Senior Consultant

London House, Inc.

Park Ridge, IL

The indifferent success in achieving the goal of reliable and valid police officer work performance measures has been attributed to the complexity of the job which includes the demands of routine patrol, unsupervised crisis intervention, and public service, and to the general inadequacy of appraisal procedures particularly when supervisory assessments constitute the main criterion. This paper explores other possible reasons for unsatisfactory police performance evaluation, reviewing and integrating previous studies and new recently collected but not yet published data. The performance criteria studied include subjective supervisory assessments (paired comparisons, behavior anchored rating scales, or traditional departmental performance ratings), and several objective criteria such as arrests, commendations and disciplinary actions.

One previous study showed that subjective ratings, both those specially designed for the research and existing departmental schemes, had considerably greater stability ($r = 0.60$) over a four-year period, than did objective indices, for which over-time correlations were generally in the 0.1 to 0.2

* Presented at the Symposium, "Psychological contributions to Police Selection and Assessemnt," 1988 APA Annual Convention, Atlant, GA.

range. A review by Froemel identified essentially three different types of relationships between the performance criteria and tenure, which would partially explain the unreliability of the objective indices. In a later study of bus operators a complicated three-way relationship between performance criteria, tenure, and race was identified. Statistical procedures were developed to identify and correct for these spurious influences thus allowing use of the total sample for validation resesarch.

The previous evidence suggested that police performance ratings should be corrected or moderated by tenure to achieve reliability, and that the correction may vary for different types of criteria.

However, earlier research did not examine the effects of correction for tenure on restriction of range on the criteria, which could lead to reduced prediction-criterion correlations. The current research involving separate job analyses at selected tenure points support and extend the implications of previous studies. Job content as described by objective indices changes as tenure inreases. We must develop the moderators or correction for the performance criteria to be able to use the total group for predictive validation purposes.

CRITERIA OF POLICE OFFICER PERFORMANCE

I. INTRODUCTION

The indifferent success in achieving the goal of reliable and valid police officer work performance measures is frequently attributed either to the complexity of the job itself (which includes the demands made by such activities as routine patrol, unsupervised crisis intervention, hazardous investigation procedures, and public service) or the general inadequacy of appraisal procedures, particularly when supervisory assessments constitute the main criterion. This paper explores other possible reasons (and statistical corrections) for inadequate police officer performance criteria. A review and integration of the results of past research led to the formulation of hypotheses concerning the causes for the evident unreliability of the performance criteria. These hypotheses were tested through the analysis of newly collected data.

II. BACKGROUND RESEARCH

From the late 1960s though 1980, the Human Resources Center of the University of Chicago was involved in a series of selection test validation research studies for police officers at municipal, state, and finally, at national levels of operation. As in almost any selection test validation, one of the most important aspects of the studies was the identification of reliable and valid criteria of performance. In all of the Human Resources Center studies, both objective and subjective performance criteria were routinely extracted and analyzed for possible use in the study. The analyses included investigations of the temporal stability of the valuables and of their interrelationships through multiple factor analyses.

In the earliest study conducted by the research team (Baehr, Furcon, & Froemel, 1968), about 160 performance measures wee screened and examined before eight were finally selected for use in the study. These eight were included a specially implemented paired comparison performance ratings as described by Baehr (1968), the 1966 departmental ratings, and six objective measures, namely, tenure, awards, citizen complaints, disciplinary actions, number of arrests made, and absences or sick days. A factor analysis of the eight measures (Baehr & Broemel, 1977) produced two factors which consisted respectively of the positive and negative indicators of police performance.

In a follow-up study in the same department, reported by Furcon, Froemel, Franczak, and Baehr (1971), the paired comparison (PC) ratings originally administered in 1966 was repeated in 1969, and measures comparable to the remaining measures were abstracted for successive years from 1966 through 1969.

A study of the measures yielded high intercorrelations for adjacent years but with sometimes rapidly diminishing correlations over a three-year time span. The subjective paired comparison rating was most stable, with a correlation of .60 over a four-year time span, followed by the subjective department rating with a correlation of .36 over a three-year span. The correlations over three years for the objective indices were: arrests .36, awards .28, citizen complaints .18, and absences .15.

Froemel (1979) reports the results of an expanded investigation of police officer performance criteria. The departmental rating and a variety of objective measures were abstracted from department files for

about 500 police officers for the years 1967 through 1972. In addition to these performance indices, there were one-time paired comparison ratings and ratings obtained from the Job Performance Description Scales (JPDS) developed by Heckman, Groner, Dunnette, and Johnson (1972). A factor analysis of all indices procedures six factors, one of which was composed of all of the subjective ratings, including the ratings from all scales of the JPDS. These results seem to indicate that both of the one-time subjective ratings produce a global measure of overall police officer performance. The objective indices, and in this instance also the departmental ratings, show the previously observed phenomenon of low stability over the six-year time span. Of the 10 indices finally retained for the study, only 2 had correlations over the six-year time span which exceeded .3 with the remainder ranging from .11 to .27.

One of the most entertaining aspects of the study was Froemel's investigation of the relationship between the objective performance criteria and tenure. The total group, which had a range for tenure from 2 to 25 years, was subdivided into 24 tenure cohorts. For each performance measures, the mean values of the criterion variable were calculated for each tenure cohort and plotted against length of service. I have taken the liberty of smoothing Froemel's original curves and eliminating values after the eighteenth year of service, since they seemed to be erratic, probably due to smaller numbers of cases.

Three different types of relationships between the performance measures and tenure seemed to emerge as a result of this procedure. The first type, shown in Figure 1, represents different patterns of a

relationship of increasing values of the performance measure by length

Insert Figure 1 about here

of service. Coversely, Figure 2 shows patterns representing a decrease in the measure by length of service. The remaining plots reported by

Insert Figure 2 about here

Froemel are more difficult to interpret. The mean number of sick days appears to be lower for low tenure groups, then increases and remains reasonably constant from 4 to 16 years of service, with a sharp increase for the highest tenure groups. There does not seem to be any discernable pattern for the mean number of misdemeanor arrests.

If we accept the fact the the curves portray stable and meaningful relationships between the objective performance index and length of service, then reading off the expected levels for the objective indices at two selected tenure levels (say 5 and 15 years of service) will provide very different descriptions of the police officer's job. In other words, the observed temporal instability of the objective indices may not be a deficiency of the index but an accurate representation of structural changes in the job itself. One can then speculate that the generally greater stability of the subjective indices is due to the fact that the assessing supervisors automatically take the varying requirements of the particular officer's job into consideration when making their assessments.

If there are real structural changes in the police officer's join that are related to varying lengths of service, then it should be

possible to detect these changes through traditional job analysis procedures. The new research undertaken for this paper investigated this hypothesis by implementing a quantitative job analysis at three different tenure points.

III. JOB ANALYSIS BY TENURE COHORTS

1. The Job Analysis Instrument

A standardized and quantified job analysis instrument was constructed in the course of a a national selection test validation study for municipal police and state highway patrol officers implemented by Furcon (1980). The Job Functions Inventory for Police Officers consists of 128 items which describe fairly specific activities performed on the job. A forced-choice response technique was used which resulted in a sort of the 128 items into four equal intervals on a scale representing the relative importance of the item for overall successful performance on the job. In the original study, eight underlying dimensions of police performance were identified.

2. The Sample

The Job Functions Inventory was completed on a volunteer basis by police officers at two mid-sized urban police departments. The usual snags and visissitudes were encountered in data collection including, in this instance rather severe time constraints and the loss of a number of incorrectly completed inventories, which resulted in the rather small sample of only 37 usable cases. However, since research with similarly constructed job description inventories for other occupations had indicated that job description profiles tended to stabilize on 10 and

sometimes even 5 sorts, it was decided to continue with the project in the hope that the results would at least be suggestive. Examination of the distribution of length-of-service for the 37 cases suggested the three length of service groupings of 0-6 years (N=12), 8 to 16 years (N=29), and 18 years or more (N=6).

3. The Study Results

A discriminant function analysis produced four bipolar functions. The two poles of each function were treated as separate factors and thus produced eight scoring dimensions for the Inventory. We were encouraged to find that, despite the small sample, each of the eight dimensions had a counterpart among the eight dimensions identified in the original study. Indeed, the correlations between similarly named dimensions ranged from .50 to .86, which testified to the stability of the structure of the instrument. Scores on each of the eight dimensions were developed for the three length-of-service groups. This is a rather surprising result in view of the small sample sizes and is probably an underestimate of the differences which actually exist among these groups. The significant tenure-related job function differences are shown in Table 1.

Insert Table 1 About Here

A possibly imaginative description of these results shows the low-tenure offices engaged in street patrol, and active and hazardous criminal investigation activities. The latter activities continue in the middle-tenure group which also undertakes special promotion and training assignments. Finally, the high-tenure group is chiefly concerned in

enforcing police procedures and regulations and carrying out such public duties as escorting V.I.P.s. With hindsight, it is now realized that the original study (Baehr, Furcon, Froemel, 1968) provided strong indications that the patrolman's job differed at different tenure points. A nodal analysis based on the eight performance measures used in the study produced eight performance patterns which were actually labeled as good or poor performance at three different tenure points: "newcomers," "established patrolmen," and "old timers."

IV. CORRECTION FOR THE EFFECT OF TENURE

If we are dealing with different "jobs" at different tenure points, then collecting objective indices for an entire police department with police officers with differing lengths of service in the course of a validation study would seem to be of dubious value. On the other hand, it might be possible to use 100% of the indices after they were corrected for the effects of tenure. A somewhat similar situation was encountered in a national selection test validation for municipal bus operators conducted by Baehr (1976) in which a rather troublesome three-way relationship between performance, tenure, and race was indentified. Low-tenure operators received lower performance indices and were also predominantly black, since they were the last ethnic group members to be hired. Under ideal circumstances, this situation could be dealt with by first correcting for race differences, then establishing the relationsip between tenure and the performance index for the "unbiased" ratings. A practical stastistical procedure for implementing these corrections, and thus making all of the ratings useable in the validation study, is reported by saunders, Baehr, and Froemel (1977).

It might well be advantageous to apply this procedure to the police officer performance indices since it is noted in both the original police patrolmen validation and the follow-up study that there are significant race differences on at least three of the eight performance criteria. White officers score higher on the departmental rating and blackk officers have higher values for number of arrests, disciplinary actions, and I.I.D. complaints. Differences on the last three indices have been thought to be due to the fact that black officers are more often assigned to tough black districts.

An initial correction for race differences may produce even clearer relationships between tenure and the performance indices. Once this relationship is established, it is a fairly simple matter to correct for tenure by subtracting the expected level of the performance measure for the particular officer's length of service from the score which he actually obtained on that measure. The resulting positive or negative score will indicate performance above or below the demands of the job for that varable at a given tenure point.

One possible difficulty with this procedure is that it may lead to a severe restriction in range of the variable concerned. It is left to future empirical research to determine whether or not the indicated corrections will produce more reliable and predictable performance criteria.

References

Baehr, M.E. (1976). National validation of a selection test battery for male transit bus operators (UMTA-MA-06-0011-77-1). Chicago: The University of Chicago, Industrial Relations Center. (NTIS No. PB-283-709/AS)

Baehr, M.E., (1968). The appraisal of job performance (Occasional Paper 27-R1). Chicago: The University of Chicago, Industrial Relations Center.

Baehr, M.E., Froemel, E.C., (1977). The arrow-dot test as a predictor of police officer performance. Perceptual and Motor Skills, 45, 683-693.

Baehr, M.E., Furcon, J.E., & Froemel, E.C. (1968). Psychological assessment of patrolman qualifications in relation to field performance. Washington, D.C.: U.S. Government Printing Office.

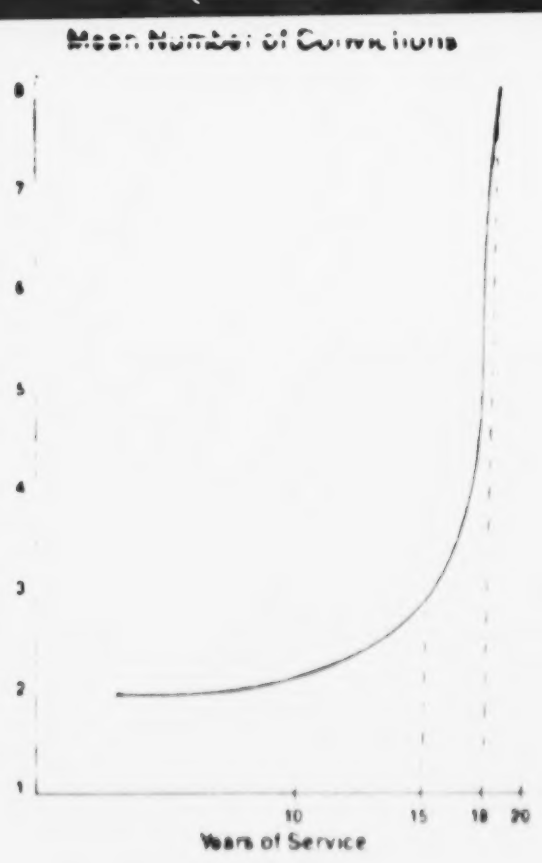
Froemel, E.C. (1979). Objective and subjective measures of police officer performance. In C. Spielberger (Ed.), Police selection and evaluation: Issues and problems. New York: Hemisphere.

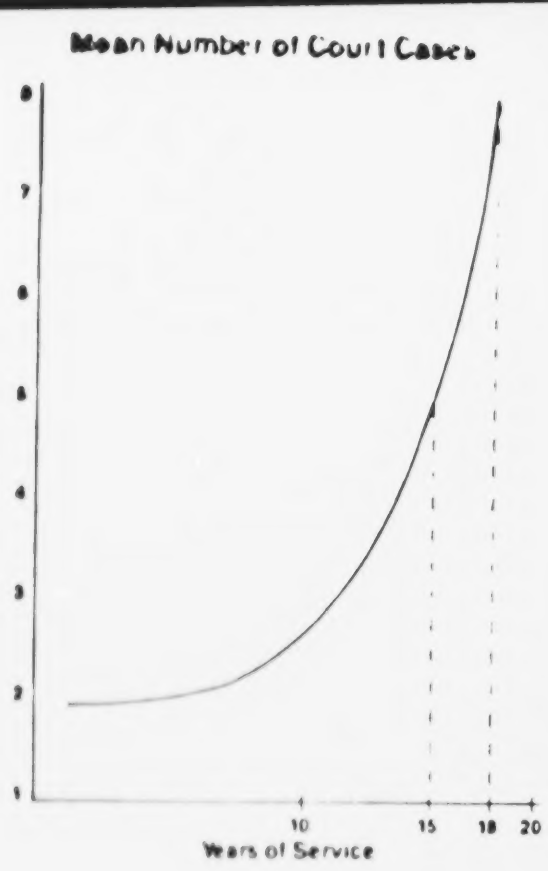
Furcon, J.E. (1980). Municipal police and state highway patrol officer selection test development project. Final Report, Chicago: The University of Chicago, Human Resource Center.

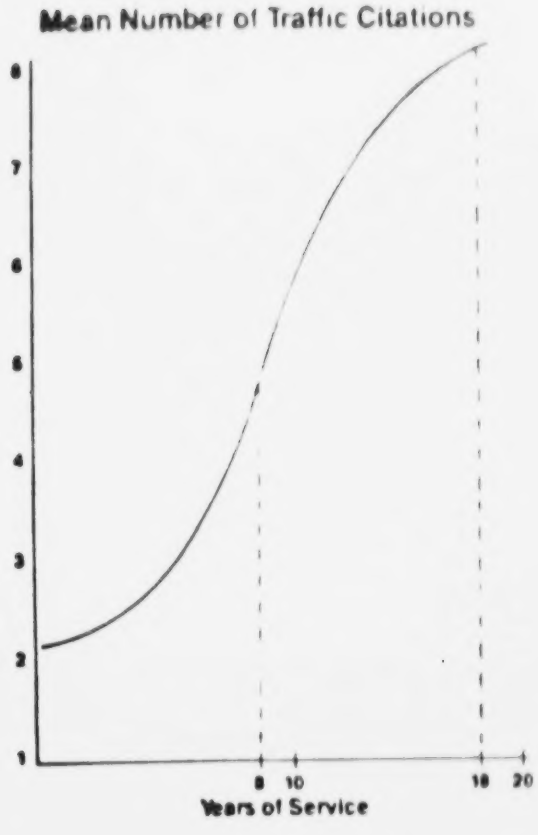
Furcon, J.E., Froemel, E.C., Franczak, R.G., & Baehr, M.E. (1971). A longitudinal study of psychological test predictors and assessments of patrolman firld performance. Chicago: The University of Chicago, Industrial Relations Center.

Heckman, R.W., Groner, D.M., Dunnette, M.D., & Johnson, P.D. (1972). Development of psychiatric standards for police selection. Minneapolis: Personnel Decisions, Inc.

Saunders, D.R., Baehr, M.E., & Froemel, E.C. (1977). Identifications of, and correction for, effects of racial bias and job tenure on supervisory ratings of bus-operator performance. Psychological Reports, 40, 859-865.







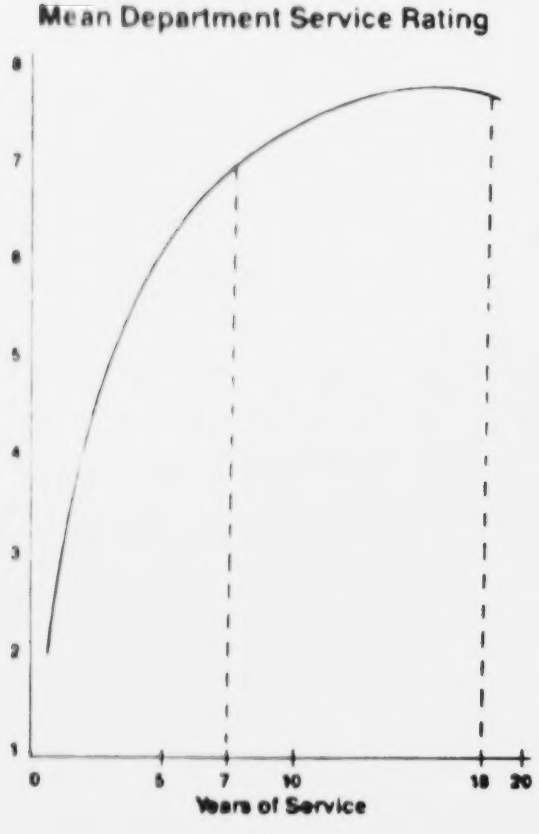


Figure 1. Patterns of increasing values of the performance measure by length of service.

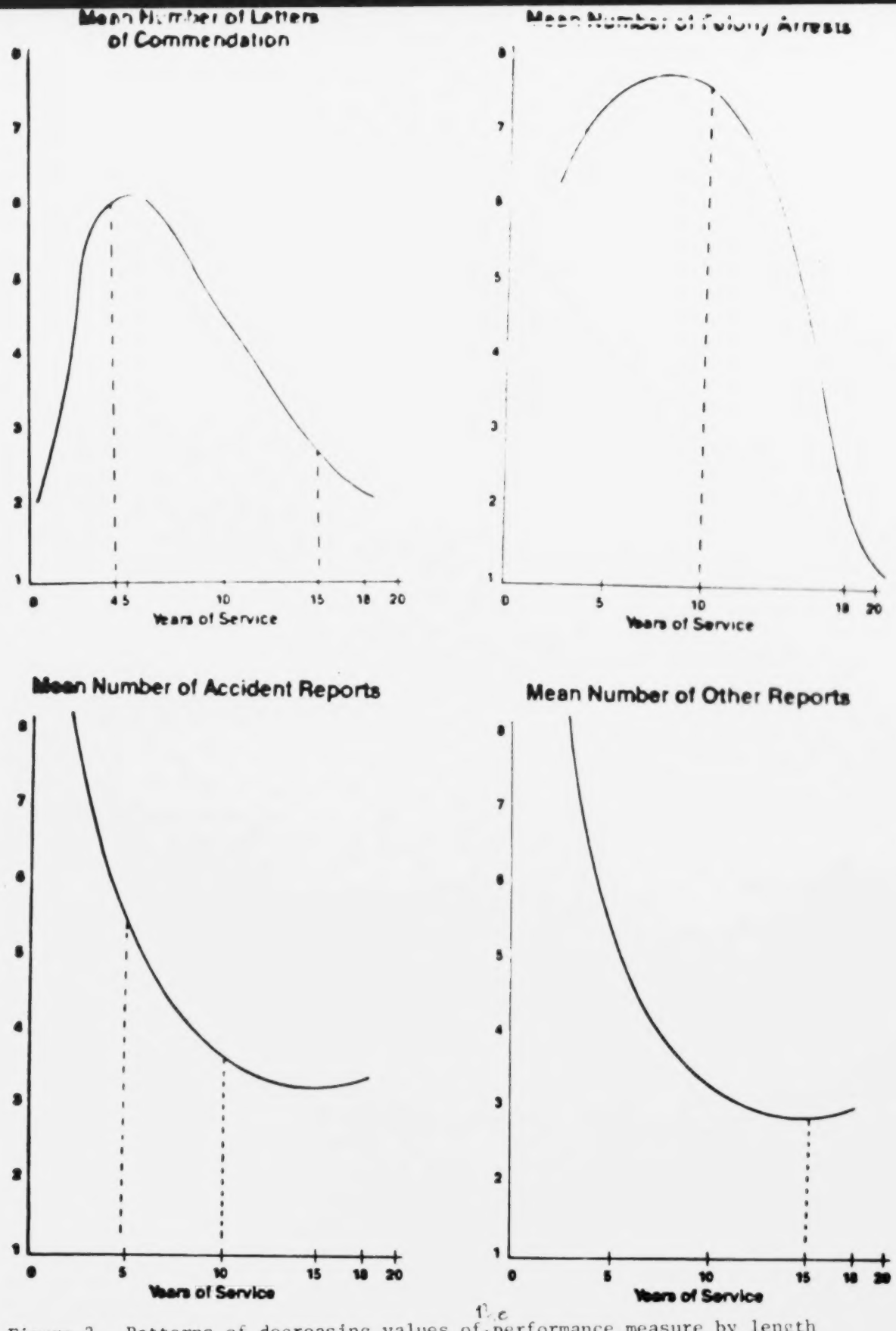


Figure 2. Patterns of decreasing values of performance measure by length of service.

Table 1. Significant Tenure-Related Job Function Differences.

6. Active/Hazardous Criminal Investigation Activities

Illustrative Items

81. Adopting civilian roles as bait in high crime areas
36. Taking part in searches for wanted criminals
82. Taking part in raids

High Group: Middle Tenure Group and Low Tenure Group
F = 3.69, p = .035

1. Public Relations/Dealing with Groups

Illustrative Items

11. Crowd Control
126. Escorting VIPs
89. Maintaining order at crime scenes

High Group: High Tenure Group and Low Tenure Group
F = 3.02, p = .062

3. Police Procedures and Regulations

Illustrative Items

46. Following proper procedures for handling prisoners
127. Following correct procedures in handling warrants
17. Serving individuals or property in accordance with regulation

High Group: High Tenure Group
F = 4.79, p = .015

7. Special Assignments/Public Promotion and Trainin

Illustrative Items

87. Training other units in special procedure
43. Serving on specially appointed task forc
91. Serving as Officer Friendly for local school

High Group: Middle Tenure Group
F = 3.29, p = .049